

REMARKS/ARGUMENTS

Applicant appreciates the allowance of claims 5, 10, 14, 17, 18, 23, 28, 32, 35, and 36.

On page 2 of the Official Action, claims 2-4, 6, 20-22, and 24 were rejected under 35 U.S.C. 103(a) as being unpatentable over Zielinski et al. (U.S. 7,487,243 B1) in view of Choquier et al. (U.S. 5,774,668). In reply, applicant respectfully traverses, and respectfully submits that one of ordinary skill in the art would not have been motivated to combine Choquier with Zielinski and to modify the combination as required to reconstruct the applicant's invention of claims 2-4, 6, 20-22, and 24.

With respect to applicant's independent claims 2 and 20, the Official Action construed the "respective utilization value of each distributed processing unit" to cover Zielinski's "resource constraints for the tunnel termination devices" such as "the maximum subscriber sessions supported by each of the tunnel termination devices ..." (See Zielinski col. 5 lines 9-18 and col. 6 lines 21.) However, applicant's claims 2 and 20 recite "the respective utilization value of said each distributed processing unit is a percentage of saturation of said each distributed processing unit." Therefore it should be clear that "resource constraints" as disclosed in Zielinski are not "utilization values." A "resource constraint" does not give any indication of the utilization of the resource, let alone a percentage of saturation of a distributed processing unit.

With respect to applicant's claims 2 and 20, page 3 paragraph 6 of the Official Action recognizes that Zielinski "does not explicitly teach wherein the respective utilization value of said each distributed processing unit is a percentage of saturation of said each distributed processing unit." The Official Action cites Choquier col. 10, lines 66 - col. 11, line 12, and col. 14, line 60 - col. 15, line 6 for teaching a respective utilization value that is a percentage of saturation of the distributed processing unit. However, neither Zielinski nor Choquier teaches applying a function to a respective percentage of saturation of each distributed processing unit to obtain a respective weight for said each distributed processing unit; and using the respective weights for the distributed processing units for distributing work requests to the distributed processing units so that the respective weight for said each distributed processing unit specifies a respective frequency at which the work requests are distributed to said each distributed processing unit. In particular, neither Zielinski nor Choquier suggests that a percentage of saturation of a distributed processing unit could or should be substituted for the resource constraint for the distributed processing unit in Zielinski.

Paragraph 7 on page 4 of the Official Action says that modifying Zielinski by incorporating the percentage of saturation of each distributed processing unit as taught by Choquier would "dynamically allocate processing resources (such as application servers) to specific on-line services, so that fluctuations in usage levels of specific on-line services can be efficiently accommodated." However, neither Zielinski nor Choquier indicates that there would be such an advantage in Zielinski. Each of Zielinski and Choquier appears to be entirely satisfactory for its intended purpose, and neither suggests the proposed combination and

modification of Zielinski so as to produce the further advantages as taught by the applicant's disclosure.

Zielinski teaches weighted load balancing among network servers that provide tunnel terminations for connecting subscriber devices, such as home computers, to a data network, such as the Internet. (Zielinski, Title, and col. 1, lines 10-21 and FIG. 1.) A network device includes a tunneling module that load balances subscriber sessions across a plurality of tunnel termination devices based on weightings associated with the tunnel termination devices. The weightings may be assigned to the tunnel termination devices by a user, or may be calculated by the network device based on resource constraints associated with the tunnel termination devices. The network device may calculate the weightings, for example, based on a maximum number of sessions supported by each of the tunnel termination devices. (Zielinski, Abstract.) As one example, LAC 10 may calculate the weightings based on a maximum number of subscriber sessions that can be supported by each of LNSs 12. Other resource constraints that may be considered include bandwidth, memory, physical location, and the like. (Zielinski, col. 3 lines 34-39; see also col. 5 lines 5-18.) In one embodiment, tunneling module 26 calculates a respective weighting for each of tunnel termination devices A-D by dividing the respective maximum subscriber sessions supported by each of the tunnel termination devices A-D by the minimum number of subscriber sessions supported on any one of the tunnel termination devices A-D. (Zielinski, col. 6, lines 4-9.)

Thus, Zielinski appears entirely satisfactory for its intended purpose of load balancing subscriber sessions across a plurality of tunnel termination devices. There is no indication in Zielinski that load balancing of fluctuations in subscriber sessions would not be efficiently

accommodated by load balancing based on resource constraints associated with the tunnel termination devices, nor is there any indication in Zielinski of fluctuations in usage levels of specific on-line services that would not be handled efficiently by handling fluctuations in subscriber sessions in the fashion disclosed in Zielinski.

Choquier discloses a system for on-line service in which gateway computers use a service map which includes application server loading conditions broadcasted by the application servers for load balancing. The Gateway microcomputers receive service requests from client microcomputers operated by end users. Upon receiving a request to open a service, the Gateway microcomputers access a periodically-updated service map to locate the replicated application servers that are currently running the corresponding service application, and then apply a load balancing method (using server load data contained within the service map) to select an application server that has a relatively low processing load. (See Choquier, title, and abstract.) Choquier can handle relatively static loads using the method illustrated in Choquier FIG. 7, and can also handle relatively dynamic loads by updating the service map more frequently and employing a randomization technique as shown in Choquier FIG. 8. (See Choquier col. 14 line 60 to col. 15 line 64.)

Thus, Choquier appears entirely satisfactory for its intended purpose of load balancing application server loads for on-line services. Choquier handles efficiently fluctuations in usage levels of specific on-line services, using methods other than the method of applicants' claims 2 and 20. .

Although both Zielinski and Choquier deal with load-balancing in a system such as the Internet in which application servers provide on-line services to subscribers or end users of client

microcomputers, Zielinski is dealing with load balancing subscriber sessions across a plurality of tunnel termination devices in the interface between the client microcomputers and the Internet, and Zielinski is not dealing with load balancing of the subscriber sessions upon the application servers that provide the on-line services. Therefore the environment and problem in Zielinski is different from the environment and problem in Choquier. There is no indication in Zielinski that relatively small changes in the number of service sessions cause significant changes in loads upon the tunnel termination devices, and instead in Zielinski it appears that the loading would be proportional to the number of service sessions. In contrast, in the environment of Choquier, for certain types of services, relatively small changes in the number of ongoing service sessions can cause significant changes in server loads, so that there could be significant fluctuations in server load independent of the number of ongoing service sessions. (See Choquier, col. 15, lines 7-14.)

In contrast to Zielinski, the applicant's disclosure teaches a dynamic form of load balancing based on various levels of usage or dynamic loading of the distributed processing units, in contrast to the relatively static weights used in Zielinski based on resource constraints of the distributed processing units. In contrast to Choquier, the applicant provides an elegant method of handling significant fluctuations in server load.

The applicant's specification "describes a method of reducing system overload conditions without the cost of additional processing units in a distributed data processing system." (Applicant's specification, page 42, lines 17-19.) The applicant's method monitors distributed performance by collecting performance statistics and does load balancing based on the collected performance statistics. The load balancing reduces dynamic overload conditions because the weighted distribution of work requests to the distributed processing units is based on a

percentage of saturation of each distributed processing unit. (See applicant's specification, page 35, lines 2-13.)

It is the percentage of saturation of each distributed processing unit that is indicative of an incipient overload condition that might be avoided by a redistribution of the incoming workload. An incipient overload condition is not indicated by a resource constraint of a distributed processing unit. The resource constraint is not indicative of an incipient overload condition that might be avoided by a redistribution of the incoming workload.

When determining whether a claim is obvious, an examiner must make "a searching comparison of the claimed invention – including all its limitations – with the teaching of the prior art." In re Ochiai, 71 F.3d 1565, 1572 (Fed. Cir. 1995) (emphasis added). Thus, "obviousness requires a suggestion of all limitations in a claim." CFMT, Inc. v. Yieldup Intern. Corp., 349 F.3d 1333, 1342 (Fed. Cir. 2003) (citing In re Royka, 490 F.2d 981, 985 (CCPA 1974)). Moreover, as the Supreme Court recently stated, "there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness." KSR Int'l v. Teleflex Inc., 127 S. Ct. 1727, 1741 (2007) (quoting In re Kahn, 441 F.3d 977, 988 (Fed. Cir. 2006) (emphasis added)). A fact finder should be aware of the distortion caused by hindsight bias and must be cautious of arguments reliant upon ex post reasoning. Id., citing Graham, 383 U.S. at 36 (warning against a "temptation to read into the prior art the teachings of the invention in issue" and instructing courts to "guard against slipping into the use of hindsight.").

Applicant's dependent claims 3-4, 6, 21-22, and 24 are patentable over the proposed combination of Zielinski and Choquier due to the limitations incorporated by reference from their respective base claims 2 and 20.

On page 5 of the Official Action, claims 12-13, 15, 30-31 and 33 were rejected under 35 U.S.C. 103(a) as being unpatentable over Zielinski et al. (U.S. 7,487,243 B1) in view of Choquier et al. (U.S. 5,774,668), and further in view of Garnett et al. (U.S. 2003/0105903 A1). In reply, applicant respectfully traverses.

With respect to the independent claims 12 and 30, Zielinski and Choquier are distinguished for the reasons discussed above with respect to applicant's claims 2 and 20. Applicant's claims 12 and 30 recite "the respective utilization value of said each virus checking server indicating a percentage of saturation of said each virus checking server." Therefore it should be clear that "resource constraints" as disclosed in Zielinski are not "utilization values." A "resource constraint" does not give any indication of the utilization of the resource, let alone a percentage of saturation of a virus checking server. Moreover, there is no indication in Zielinski that the tunnel termination devices are virus checking servers. Nor is there any disclosure of virus checking in Choquier or Garnett.

Each of Zielinski, Choquier and Garnett appears entirely satisfactory for its intended purpose, so that the references as a whole do not suggest that Zielinsky should be combined with Choquier and Garnet and modified in the fashion proposed in the Official Action. Thus, Zielinski appears entirely satisfactory for its intended purpose of load balancing subscriber sessions across a plurality of tunnel termination devices. There is no indication in Zielinski that

load balancing of fluctuations in subscriber sessions would not be efficiently accommodated by load balancing based on resource constraints associated with the tunnel termination devices, nor is there any indication in Zielinski of fluctuations in usage levels of specific on-line services that would not be handled efficiently by handling fluctuations in subscriber sessions in the fashion disclosed in Zielinski. In a similar fashion, Choquier appears entirely satisfactory for its intended purpose of load balancing application server loads for on-line services. The environment and problem in Zielinski is different from the environment and problem in Choquier. Zielinski is dealing with load balancing subscriber sessions across a plurality of tunnel termination devices in the interface between the client microcomputers and the Internet, and Zielinski is not dealing with load balancing of the subscriber sessions upon the application servers that provide the on-line services. There is no indication in Zielinski that relatively small changes in the number of service sessions cause significant changes in loads upon the tunnel termination devices, and instead in Zielinski it appears that the loading would be proportional to the number of service sessions. In contrast, in the environment of Choquier, for certain types of services, relatively small changes in the number of ongoing service sessions can cause significant changes in server loads, so that there could be significant fluctuations in server load independent of the number of ongoing service sessions.

Garnett is cited on page 6 of the Official Action for disclosing round-robin load balancing, paragraphs 216-218. Garnett also appears entirely satisfactory for its intended purpose and does not provide motivation for modifying Zielinski in view of Choquier in order to reconstruct the invention of applicant's claims 12 and 30. Garnett deals with a server blade performing load balancing functions and describes various load balancing algorithms for

distributing new connections among servers in a group of servers. (See Garnett, title, and paragraph [0216] to [0220].) Garnett does not specifically deal with the problem of load balancing subscriber sessions across a plurality of tunnel termination devices as in Zielinski, and Garnett does not suggest that Zielinski should be modified to obtain a respective utilization value of each tunnel termination device, the respective utilization value indicating a percentage of saturation of each tunnel termination device, and to apply a function to the respective utilization value to obtain a respective weight for each tunnel termination device.

Applicant's dependent claims 13, 15, 31, and 33 are distinguished from the proposed combination of Zielinski, Choquier, and Garnett due to the limitations incorporated by reference from their respective base claims 12 and 30.

On page 7 of the Official Action, claims 16 and 34 were rejected under 35 U.S.C. 103(a) as being unpatentable over Zielinski et al. (U.S. 7,487,243 B1) in view of Choquier et al. (U.S. 5,774,668), and further in view of Garnett et al. (U.S. 2003/0105903 A1, as applied to claims 12 and 30 above, and further in view of Kapoor (U.S. 5,884,038). In reply, applicant respectfully traverses and submits that claims 16 and 35 are patentable due to the limitations incorporated by reference from their base claims 12 and 30, because Kapoor does not disclose or suggest the limitations of the base claims 12 and 30 that are missing from Zielinski, Choquier, and Garnett. Kapoor (col. 5 lines 11-16 and 60-63) was cited for accessing a randomized distribution list for distributing work requests to distributed processing units as indicated by entries in the randomized distribution list. Kapoor does not specifically deal with the problem of load balancing subscriber sessions across a plurality of tunnel termination devices as in Zielinski, and

Kapoor does not suggest that Zielinski should be modified to obtain a respective utilization value of each tunnel termination device, the respective utilization value indicating a percentage of saturation of each tunnel termination device, and to apply a function to the respective utilization value to obtain a respective weight for each tunnel termination device.

In view of the above, it is respectfully submitted that the application is in condition for allowance. Reconsideration and early allowance are earnestly solicited.

Respectfully submitted,

/ *Richard C. Auchterlonie* /

Richard C. Auchterlonie, Reg. No. 30,607

NOVAK DRUCE & QUIGG, LLP
1000 Louisiana, 53rd Floor
Houston, TX 77002
713-571-3460 (Phone)
713-456-2836 (Fax)
Richard.Auchterlonie@novakdruce.com